Exam1: Introduction to Probability Modeling, 36-410 (Spring 2017)

Name: _______________________

Andrew ID: ___________________

<table>
<thead>
<tr>
<th>Question #</th>
<th>Points Possible</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11 + 2 extra credit</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5 extra credit</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35 + 7 extra credit</td>
<td></td>
</tr>
</tbody>
</table>

Extra credit only helps if you have a total score less than 35, i.e. your final score will be min (score + extra credit, 35). In general, do not assume extra credit questions are more difficult than the others. They are chosen close to randomly.
Instructions:

• You have 70 minutes to work on the exam. Please understand that in order to be fair, we will stop at 1:15, regardless of the time that you started the exam.

• Illegible work will not receive any credit.

• In general, there will not be any partial credit for these questions.

• If you provide multiple answers, the question will be marked incorrect.

• No use of calculators or notes during the exam.

• There are no inverses provided this time. If you need a matrix inverse (and you should) then write out clearly what the matrix you need to invert is and what your subsequent operations are. For example you can say:

  “To obtain the answer invert the matrix

  \[
  M = \begin{bmatrix}
  0.4 & 0.4 & 0.2 \\
  0 & 0 & 0 \\
  0 & 0.1 & 0.2
  \end{bmatrix},
  \]

  and compute the sum of the first row.”

  You have to write out exactly and explicitly what the matrix is, i.e. you cannot say invert the transient block of the transition matrix, and be extremely clear about what the subsequent operations are.

• There is lots of scratch paper in the front of the room. Raise your hand if you need some.

• If you need more space to explain your answer, get some scratch paper from me, and staple the scratch paper to your exam when turning it in. I have a stapler.

• I prefer if you answer the exam in pen. If you choose to answer in pencil that is fine, except I will not accept re-grade requests unless they happen immediately after I hand back your exams.
1. [5 points] Virus Mutations: Suppose that a virus can exist in one of \( N \) different strains. In each generation, the virus either stays the same or with probability \( \alpha \) mutates to another strain (chosen uniformly at random from the remaining strains).

We are interested in the probability that the strain in the \( n \)th generation is the same as in the 0th, i.e. we are interested in \( P_{ii}^{(n)} \) for each \( i \). Show that this probability satisfies the recursion for \( n \geq 2 \):

\[
P_{ii}^{(n)} = \frac{\alpha}{N - 1} + \left(1 - \frac{N\alpha}{N - 1}\right) P_{ii}^{(n-1)}.
\]

Hint: There are many ways to solve this problem but try to think of the simplest Markov chain that suffices for your calculation.

Solution: The easiest way to solve this problem is to notice that we only need to consider a Markov chain with two states \( \{0, 1\} \) where \( X_n = 0 \) if the virus is in strain \( i \), and \( X_n = 1 \) otherwise. The transition matrix is:

\[
P = \begin{bmatrix}
1 - \alpha & \alpha \\
\frac{\alpha}{N-1} & 1 - \frac{\alpha}{N-1}
\end{bmatrix}.
\]

Suppose, that we were given the \( n - 1 \)-step transition matrix, i.e. for some values \( \theta_1, \theta_2 \) it takes the form:

\[
P^{(n-1)} = \begin{bmatrix}
\theta_1 & 1 - \theta_1 \\
\theta_2 & 1 - \theta_2
\end{bmatrix}.
\]

Then we could compute the \( n \)-step transition matrix using the C-K identity as:

\[
P^{(n)} = P^{(n-1)} P.
\]

We are interested in finding, \( P_{11}^{(n)} \) (recall that we mapped the \( i \)th state to 0). So have that:

\[
P_{11}^{(n)} = (1 - \alpha)\theta_1 + (1 - \theta_1) \frac{\alpha}{N - 1}
\]

\[
= \frac{\alpha}{N - 1} + \left(1 - \frac{N\alpha}{N - 1}\right) \theta_1.
\]

Observing that \( \theta_1 = P_{11}^{(n-1)} \) and once again recalling that we mapped the \( i \)th state to 0, gives us the answer.
2. [9 points] Queuing Theory: An important application of Markov chains is in modelling queueing systems. Consider the following system, there is queue of length 2, i.e. there can be 0, 1 or 2 jobs in the queue. Each day one of 3 things happens, either a new job arrives (with probability $p$), or an existing job (if there is one) is completed (with probability $q$), or nothing happens.

If a new job arrives when the queue is full, it is simply rejected.

- [2 points] Set this up as a Markov chain, where the state space is the number of jobs currently in the queue. What is the transition matrix?
  
  **Solution:** The transition matrix is:
  
  $$
  P = \begin{bmatrix}
  1 - p & p & 0 \\
  q & 1 - p - q & p \\
  0 & q & 1 - q 
  \end{bmatrix}.
  $$

- [1 point] What are the classes of this Markov chain. Is this Markov chain, irreducible, aperiodic?
  
  **Solution:** The Markov chain is irreducible, i.e. all states are in a single class, and aperiodic.
• [4 points] Calculate the limiting distribution of the Markov chain. This should be a function of \( p \) and \( q \). **Hint:** Set up the appropriate system of equations, and solve it.

**Solution:** We set up the system of equations. As is always the case we can ignore one of the equations (since it is a redundant system) so we choose the following:

\[
\begin{align*}
\pi_0 + \pi_1 + \pi_2 &= 1 \\
(1 - p)\pi_0 + q\pi_1 &= \pi_0 \implies q\pi_1 = p\pi_0 \\
p\pi_1 + (1 - q)\pi_2 &= \pi_2 \implies q\pi_2 = p\pi_1.
\end{align*}
\]

Substituting back into the first equation we obtain:

\[
\begin{align*}
\pi_0 &= \frac{1}{(1 + p/q + p^2/q^2)} \\
\pi_1 &= \frac{p/q}{(1 + p/q + p^2/q^2)} \\
\pi_2 &= \frac{p^2/q^2}{(1 + p/q + p^2/q^2)}.
\end{align*}
\]

This is the desired limiting distribution.

• [2 points] In the long run, what fraction of jobs will be rejected?

**Solution:** In the long-run we reject any job that arrives when we are in State 2, i.e. we reject a job with probability

\[
p\pi_2 = \frac{p^3/q^2}{(1 + p/q + p^2/q^2)}.
\]
3. We have a Markov chain with state space \( S = \{0, 1, 2, \ldots, 12\} \), with transition matrix given below: note that the states are numbered from 0. I have tried to make your work (slightly) easier since for this MC classes are groups of adjacent states.

\[
P = \begin{bmatrix}
0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0
\end{bmatrix}
\]

\[
P^{(10000)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0.2143 & 0.4286 & 0.2143 & 0.0476 & 0.0476 & 0.0476 & 0.0476 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1786 & 0.3571 & 0.1786 & 0.0952 & 0.0952 & 0.0952 & 0.0952 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1786 & 0.3571 & 0.1786 & 0.0952 & 0.0952 & 0.0952 & 0.0952 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.1429 & 0.2857 & 0.1429 & 0.1429 & 0.1429 & 0.1429 & 0.1429 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2143 & 0.4286 & 0.2143 & 0.0476 & 0.0476 & 0.0476 & 0.0476 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2143 & 0.4286 & 0.2143 & 0.0476 & 0.0476 & 0.0476 & 0.0476 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2143 & 0.4286 & 0.2143 & 0.0476 & 0.0476 & 0.0476 & 0.0476 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2143 & 0.4286 & 0.2143 & 0.0476 & 0.0476 & 0.0476 & 0.0476 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0.3333 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0.3333 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3333 & 0.3333 & 0.3333 & 0.3333
\end{bmatrix}
\]
(a) [2 points] Neatly draw the state diagram for this chain.
(b) [1 point] Find the communicating classes for this chain. Solution: The communicating classes are:

- \{0, 1, 2\}
- \{3, 4, 5, 6\}
- \{7, 8, 9\}
- \{10, 11, 12\}

(c) [1 point] For each communicating class indicate whether it is recurrent or transient.

Solution:

- \{0, 1, 2\} – Transient
- \{3, 4, 5, 6\} – Transient
- \{7, 8, 9\} – Recurrent
- \{10, 11, 12\} – Recurrent

(d) [1 point] For each communicating class, give its period.

Solution: The periods are:

- \{0, 1, 2\} – 3
- \{3, 4, 5, 6\} – 3
- \{7, 8, 9\} – Aperiodic
- \{10, 11, 12\} – Aperiodic

(e) [2 points] Assuming the chain starts in state 4, find the expected time until the chain hits an absorbing class.

Solution: We can find the matrix \( T \) as:

\[
T = \begin{bmatrix}
0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1/3 & 0 & 0 & 0
\end{bmatrix}.
\]

We can then calculate the inverse:

\[
U = (I - T)^{-1} = \begin{bmatrix}
1 & -1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 \\
-1/2 & 0 & 1 & -1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1/3 & -1/3 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & -1/3 & 0 & 0 & 1
\end{bmatrix}.
\]

and sum the entries in the fourth row (starting from 0).
(f) [1 point] Assuming the chain starts in state 10, find the long-run proportion of time the chain spends in state 11.

**Solution:** We look at the corresponding entry of $P^{(10000)}$ to see that this is $1/3$.

(g) [1 point] Assuming the chain starts in state 7, find the long-run proportion of time the chain spends in state 8.

**Solution:** Again we look at the corresponding entry of $P^{(10000)}$ to see that this is $0.5$. 
(h) [2 points] Describe in detail the long-term behavior of this chain.

Solution: The chain has two transient and two recurrent classes. Both the recurrent classes are reachable from either of the transient classes, so if we begin in the transient classes \{0, 1, 2\} or \{3, 4, 5, 6\} we end up in the one of the recurrent classes (the long run probabilities are in the matrix $P^{(10000)}$).

On the other hand if we start in the class \{7, 8, 9\} we remain in this class in the long run and converge to the distribution \[0.250, 0.50, 0.25\] on these three states. Similarly, if we start in the class \{10, 11, 12\} we remain in this class and converge to the uniform distribution on these states.

(i) [2 points] Extra Credit Suppose that we begin in the transient state 4, what is the probability of being absorbed into the recurrent class containing the state 8?

Solution: We simply sum the entries of the matrix $P^{10000}$ to obtain the probability $= 0.1429 \times 2 + 0.2857 = 0.5715$. 


4. **[5 points] Motif Finding:** An important problem in genetics, is to identify frequently occurring patterns in our DNA (these are known as motifs). Suppose that our DNA is a sequence made up of 4 characters: \( \{A,G,T,C\} \). Further, for simplicity assume that they are all equally likely and random, i.e. every subsequent character is *independently* one of the above 4 characters.

Suppose, we are interested in the motif ‘GAT’, i.e. we are interested in how often this sequence occurs, if the sequence was genuinely random.

- **[2 points]** Set up the above process as a Markov chain. Particularly, describe an appropriate state space (see the next question to get some idea of what appropriate means), the initial state, and the transition matrix.

**Solution:** We let the state space be \( X_n = \{0,1,2,3\} \), where \( X_n = 0 \) if we matched none of the desired motif so far, \( X_n = 1 \) if we have matched the letter ‘G’ and so on. Once we match the full sequence, we remain in the state 3.

The transition matrix is:

\[
P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\]

- **[3 points]** Calculate the expected length of the sequence before you hit the desired motif.

**Solution:** This is just the expected time spent in the transient states. We compute the fundamental matrix as:

\[
U = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & 0 \\ -\frac{1}{2} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{2} & -\frac{1}{4} & 1 \end{bmatrix}^{-1},
\]

and sum the first row.
5. [5 points] Coupon Collector: A classical problem that arises in many different contexts is known as the coupon collector problem. There are $n$ different types of coupons, and a coupon collector wants to collect all $n$ types. On the $0^{th}$ day she has no coupons. Every day after that, she is given a coupon uniformly at random from each of the $n$ types, i.e. she is equally likely to receive a coupon of any type. She stops when she receives all the different types of coupons.

- [2 points] Set up the above process as a Markov chain. Particularly, describe an appropriate state space (see the next question to get some idea of what appropriate means), the initial state, and the transition matrix.

**Solution:** In this case since we want to calculate the expected time to collect all the different coupons we can use the state space as $\{0, 1, \ldots, n\}$, denoting the number of different coupons she has collected so far. The initial state is $X_0 = 0$, and the transition matrix is:

$$P = \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 & 0 \\
0 & \frac{1}{n} & \frac{n-1}{n} & 0 & \ldots & 0 \\
0 & 0 & \frac{2}{n} & \frac{n-2}{n} & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 1
\end{bmatrix}.$$

- [3 points] Take $n = 3$. Calculate the expected number of days until she collects all the different types of coupons.

**Solution:** When $n = 3$ we have the transition matrix:

$$P_3 = \begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & \frac{1}{3} & \frac{2}{3} & 0 \\
0 & 0 & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & 1
\end{bmatrix}.$$

We want to calculate the time spent in transient states so we compute the fundamental matrix:

$$U = \begin{bmatrix}
1 & -1 & 0 \\
0 & \frac{2}{3} & -\frac{2}{3} \\
0 & 0 & \frac{2}{3}
\end{bmatrix}^{-1},$$

and sum the entries of the first row.
6. (a) [3 points] Extra Credit Suppose that \( \{X_n\}_{n \geq 0} \) is a first-order time-homogenous Markov chain. Consider the new stochastic process \( \{Y_n\}_{n \geq 0} \), where each
\[
Y_n = X_{kn},
\]
where \( k > 1 \) is a positive integer.

- Is this stochastic process a Markov chain? If yes, prove the Markov property. If no, explain why not.

**Solution:** It is a Markov chain. To verify the Markov property we note that for any \( n \),
\[
Y_{n-1} \perp \perp Y_{n+1}|Y_n.
\]
Observe that the above claim is just that:
\[
X_{kn-k} \perp \perp X_{kn+k}|X_{kn}.
\]
This follows simply from the Markov property for \( \{X_n\} \).

- If it is a Markov chain, what is its order? (Your answer can depend on \( k \)).

**Solution:** It is a first-order Markov chain, since as we showed above \( Y_{n-1} \perp \perp Y_{n+1}|Y_n \).

- If it is a Markov chain, is it time-homogenous? What is its transition matrix?

**Solution:** It is time-homogenous. The transition matrix for the \( Y \) process is simply \( P^k \), where \( P \) is the transition matrix of the first-order chain \( \{X_n\} \). To see this observe that for any pair of states \( i, j \):
\[
P(Y_{n+1} = j|Y_n = i) = P(X_{kn+k} = j|X_{kn} = i) = P_{ij}^{(k)}.
\]
(b) [2 points] Extra credit Recall, the PageRank algorithm from lecture. We created a new transition matrix of the form:

\[ Q = (1 - \theta)E/n + \theta P, \]

where \( E \) is the matrix of all ones. The addition of the matrix \( E \) ensured that the Markov chain had a unique limiting distribution. We said that a typical choice of \( \theta = 0.85 \). Of course, the Markov chain has a unique limiting distribution even if \( \theta = 0.9999 \), and this weights the true web-graph more strongly (which is desirable).

However, setting \( \theta \) very close to 1 has algorithmic implications. The algorithm we discussed in lecture of starting from some initial state and then simply running the Markov chain can be very slow to converge to the limiting distribution.

Consider a simple webgraph, like the one shown in the figure. Intuitively, explain what happens to the PageRank algorithm if we take \( \theta \) very close to 1.

**Solution:** The limiting distribution in this case is uniform. However, suppose we initialize in State 0, and run the PageRank algorithm, then it will be very slow to converge. The distribution of \( \pi^{(t)} \) will be quite close to the uniform distribution on the first group \{0, 1, 2\} and 0 on the second group, and will slowly adjust to the uniform distribution on all nodes \{0, 1, 2, 3, 4, 5\}.

In order to ensure speedy convergence we need to take \( \theta \) to not be too close to 1, and this gives some justification for the practical choice of selecting \( \theta = 0.85 \).